

## **Proyecto de ETL: Análisis de empresas del S&P 500**

### **Objetivo:**

Desarrollar un proceso ETL automatizado para extraer la lista de empresas del S&P 500 desde Wikipedia, obtener información y descargar los precios diarios de cotización del último trimestre. Los estudiantes deberán implementar funciones para las fases de extracción (EXTRACT), transformación (TRANSFORM) y carga (LOAD) de los datos.

### **Requisitos:**

- Python 3.x
- Librerías: **requests**, **BeautifulSoup**, **pandas**, **yfinance**, **logging**

### **Instrucciones Generales:**

1. Configura un entorno virtual e instala las librerías necesarias.
2. Implementa el proceso ETL en Python, organizando el código en funciones claras para cada fase (EXTRACT, TRANSFORM, LOAD).
3. Utiliza logging para registrar el progreso y cualquier error que ocurra durante el proceso.

### **Fases del Proyecto ETL:**

#### **1. Configuración del Entorno**

##### **Pasos a seguir:**

1. Crear un entorno virtual.
2. Instalar las librerías necesarias utilizando **pip**.

#### **2. Extracción de Datos (EXTRACT)**

**Objetivo:** Obtener la lista de empresas del S&P 500 desde Wikipedia y los perfiles de cada empresa desde Yahoo Finance.

##### **Tareas:**

### 1. Extraer la lista de empresas del S&P 500 desde Wikipedia:

- Usar **requests** para hacer una solicitud HTTP a la página de Wikipedia ([https://es.wikipedia.org/wiki/Anexo:Compa%C3%B1%C3%ADas\\_del\\_S%26P\\_500](https://es.wikipedia.org/wiki/Anexo:Compa%C3%B1%C3%ADas_del_S%26P_500)).
- Usar **BeautifulSoup** para parsear el HTML y extraer la tabla de empresas.
- Guardar los datos en un DataFrame de **pandas**.

### 2. Obtener los precios de cada empresa desde Yahoo Finance:

- Utilizar la librería **yfinance** para extraer información de los precios de cotización (Close) de cada empresa (Ultimo trimestre).
- Guardar los perfiles en un DataFrame de **pandas**.

## 3. Transformación de Datos (TRANSFORM)

**Objetivo:** Limpiar y transformar los datos extraídos para que sean consistentes y adecuados para el análisis.

### Tareas:

#### 1. Limpiar los datos de la lista de empresas:

- Eliminar columnas innecesarias.
- Renombrar columnas para mayor claridad.
- Manejar valores faltantes (NaN).

#### 2. Limpiar los datos de los precios de las empresas:

- Seleccionar solo las columnas relevantes (Date, Ticker, Close).
- Estandarizar los formatos de datos (e.g., fechas, números).

## 4. Carga de Datos (LOAD)

**Objetivo:** Guardar los datos transformados en archivos CSV para su posterior análisis.

**Tareas:**

1. **Guardar la lista de empresas transformada en un archivo CSV.**
2. **Guardar los precios de las empresas transformados en un archivo CSV.**

**Resumen:**

Los estudiantes deberán implementar funciones para cada fase (EXTRACT, TRANSFORM, LOAD) y utilizar logging para monitorear el proceso.

Al finalizar, deberían tener un conjunto de datos limpio y estructurado para su posterior análisis y modelado.